

# 第一章

## 条件期望、条件方差和线性投影

# Outline

- 1 条件期望
- 2 条件方差
- 3 线性投影

# 条件期望的定义

令  $\mathbf{x}$  为随机向量,  $y$  为一元随机变量。以  $\mathbf{x}$  为条件,  $y$  的条件均值函数 (Conditional Expectation Function, CEF) 定义如下：

$$\mu(\mathbf{x}) = \mathbb{E}(y|\mathbf{x}) = \begin{cases} \int_{\mathcal{Y}} y \cdot f(y|\mathbf{x}) dy, & y \text{ 连续}, \\ \sum_{\mathcal{Y}} y \cdot P_{y|\mathbf{x}}(y|\mathbf{x}), & y \text{ 离散}, \end{cases}$$

其中,  $f(y|\mathbf{x})$  是条件概率密度函数,  $P_{y|\mathbf{x}}(y|\mathbf{x})$  是条件概率。

# 条件期望的性质

迭代期望律 ( Law of Iterated Expectations, LIE )

$$\mathbb{E}_{\mathbf{x}} [\mathbb{E}[y|\mathbf{x}]] = \mathbb{E}[y]$$

Proof.

$$\begin{aligned}\mathbb{E}_{\mathbf{x}} [\mathbb{E}[y|\mathbf{x}]] &= \int_{\mathbf{x}} \left( \int_y y \cdot f(y|x) dy \right) f(x) dx \\ &= \int_{\mathbf{x}} \int_y y f(x, y) dx dy \\ &= \int_y y f(y) dy = \mathbb{E}[y].\end{aligned}$$



## 一般形式的迭代期望律

令  $\mathbf{w}$  为随机向量,  $y$  是一元随机变量,  $\mathbf{x}$  是随机向量且满足  $\mathbf{x} = \mathbf{f}(\mathbf{w})$ .<sup>a</sup> 一般形式的 LIE 可表示为:

$$\mathbb{E}(y|\mathbf{x}) = \mathbb{E}[\mathbb{E}(y|\mathbf{w})|\mathbf{x}].^b$$

“The smaller information set always dominates.”

特例:  $\mathbb{E}(y|\mathbf{x}) = \mathbb{E}[\mathbb{E}(y|\mathbf{x}, \mathbf{z})|\mathbf{x}]$ .

<sup>a</sup> $\mathbf{x} = \mathbf{f}(\mathbf{w})$  implies that if we know the outcome of  $\mathbf{w}$ , then we know the outcome of  $\mathbf{x}$ .

<sup>b</sup>Note that we also have  $\mathbb{E}(y|\mathbf{x}) = \mathbb{E}[\mathbb{E}(y|\mathbf{x})|\mathbf{w}]$ .

## 条件期望的线性性质

$$\mathbb{E}\left(\sum_{j=1}^G a_j(\mathbf{x})y_j + b(\mathbf{x})|\mathbf{x}\right) = \sum_{j=1}^G a_j(\mathbf{x})\mathbb{E}(y_j|\mathbf{x}) + b(\mathbf{x}).$$

## The CEF Decomposition Property

For any random variable  $y$ , we have

$$y = \mathbb{E}(y|\mathbf{x}) + \varepsilon,$$

where

- $\varepsilon$  is mean independent of  $\mathbf{x}$ , that is,  $\mathbb{E}(\varepsilon|\mathbf{x}) = 0$ ,
- $\varepsilon$  is uncorrelated with any function of  $\mathbf{x}$ .

Proof.

First, note that

$$\mathbb{E}(\varepsilon|\mathbf{x}) = \mathbb{E}[y - \mathbb{E}(y|\mathbf{x})|\mathbf{x}] = \mathbb{E}(y|\mathbf{x}) - \mathbb{E}(y|\mathbf{x}) = 0.$$

Next, by property LIE, for any function  $g(\cdot)$ ,

$$\mathbb{E}[g(\mathbf{x})\varepsilon] = \mathbb{E}(\mathbb{E}[g(\mathbf{x})\varepsilon|\mathbf{x}]) = \mathbb{E}(g(\mathbf{x})\mathbb{E}[\varepsilon|\mathbf{x}]) = 0.$$



## The CEF Prediction Property

If  $\mathbb{E}(y^2) < \infty$  and  $\mu(\mathbf{x}) \equiv \mathbb{E}(y|\mathbf{x})$ , then  $\mu$  is a solution to

$$\min_{m \in \mathcal{M}} \mathbb{E}[(y - m(\mathbf{x}))^2],$$

where  $\mathcal{M}$  is the set of functions  $m : \mathbb{R}^K \rightarrow \mathbb{R}$  such that  $\mathbb{E}[m(\mathbf{x})^2] < \infty$ . In other words,  $\mu(\mathbf{x})$  is the best mean square predictor of  $y$  based on information contained in  $\mathbf{x}$ .



# 条件方差

以  $\mathbf{x}$  为条件,  $y$  的条件方差定义为：

$$\text{Var}(y|\mathbf{x}) \equiv \sigma^2(\mathbf{x}) \equiv \mathbb{E}[\{y - \mathbb{E}(y|\mathbf{x})\}^2|\mathbf{x}] = \mathbb{E}(y^2|\mathbf{x}) - [\mathbb{E}(y|\mathbf{x})]^2.$$

根据条件方差的定义，容易证明：

$$\text{Var}[a(\mathbf{x})y + b(\mathbf{x})|\mathbf{x}] = [a(\mathbf{x})]^2 \text{Var}(y|\mathbf{x}).$$

## The ANOVA Theorem

$$\text{Var}(y) = \mathbb{E}[\text{Var}(y|\mathbf{x})] + \text{Var}[\mathbb{E}(y|\mathbf{x})].$$

### Proof.

The CEF decomposition property implies that

$$\text{Var}(y) = \text{Var}[\mathbb{E}(y|\mathbf{x})] + \text{Var}(\varepsilon).$$

And, note that

$$\text{Var}(\varepsilon) = \mathbb{E}(\varepsilon^2) = \mathbb{E}[\mathbb{E}(\varepsilon^2|\mathbf{x})] = \mathbb{E}[\text{Var}(y|\mathbf{x})].$$



# 线性投影

- 由 CEF Prediction Property 定理可知  $\mu(\mathbf{x}) \equiv \mathbb{E}(y|\mathbf{x})$  是 best mean square predictor of  $y$ .
- $\mu(\mathbf{x})$  一般是未知的, 在实证分析中, 通常采用线性函数近似 CEF:

$$\mu(\mathbf{x}) = \mathbf{x}'\boldsymbol{\beta}$$

## Best Linear Predictor

The best linear predictor of  $y$  given  $\mathbf{x}$  is

$$\mathbb{L}(y|\mathbf{x}) = \mathbf{x}'\boldsymbol{\beta}$$

where  $\boldsymbol{\beta}$  minimizes the mean squared prediction error

$$S(\boldsymbol{\beta}) = \mathbb{E} \left( (y - \mathbf{x}'\boldsymbol{\beta})^2 \right).$$

The minimizer

$$\boldsymbol{\beta} = \arg \min_{\mathbf{b} \in \mathbb{R}^K} S(\mathbf{b})$$

is called the **Linear Projection (线性投影) Coefficient**.

定义  $e = y - \mathbf{x}'\beta$ , 由最优化的一阶条件, 我们有 :

$$\mathbb{E}(\mathbf{x}e) = \mathbf{0}.$$

于是,  $\beta = (\mathbb{E}(\mathbf{x}\mathbf{x}'))^{-1} \mathbb{E}(\mathbf{x}y)$ .

- 如果  $\mathbf{x} = (1, x_1)'$ , 则  $\beta_1 = \frac{\text{Cov}(y, x_1)}{\text{Var}(x_1)}$ ,  $\alpha = \mathbb{E}(y) - \beta_1 \mathbb{E}(x_1)$ .
- 如果  $\mathbf{x} = (1, x_1, \dots, x_K)'$ , 其中  $K > 1$ , 则

$$\beta_k = \frac{\text{Cov}(y, \tilde{x}_k)}{\text{Var}(\tilde{x}_k)}, \quad (1)$$

其中,  $\tilde{x}_k$  是  $x_k$  对所有其他协变量投影得到的残差。

- It shows us that each coefficient in a multivariate regression is the bivariate slope coefficient for the corresponding regressor after partialing out all the other covariates.

To verify (1), 将

$$y = \alpha + \beta_1 x_1 + \cdots + \beta_k x_k + \cdots + \beta_K x_K + e$$

代入  $\text{Cov}(y, \tilde{x}_k)$ , 注意到 :

- $\text{Cov}(\tilde{x}_k, e) = 0$ ,
- $\text{Cov}(\tilde{x}_k, x_j) = 0, j \neq k$ ,
- $\text{Cov}(\tilde{x}_k, x_k) = \text{Cov}(\tilde{x}_k, \tilde{x}_k + \mathbf{x}'_{-k}\boldsymbol{\beta}_{-k}) = \text{Var}(\tilde{x}_k)$ .<sup>1</sup>

于是,

$$\frac{\text{Cov}(y, \tilde{x}_k)}{\text{Var}(\tilde{x}_k)} = \frac{\beta_k \text{Var}(\tilde{x}_k)}{\text{Var}(\tilde{x}_k)} = \beta_k.$$

<sup>1</sup> $\mathbf{x}_{-k}$  表示除  $x_k$  之外的所有协变量构成的向量。

Below we discuss three reasons why the vector of Linear Projection Coefficients might be of interest.

### The Linear CEF Theorem (Regression Justification I)

Suppose the CEF is linear. Then  $\mathbb{L}(y|\mathbf{x})$  is it.

Proof.

假设  $\mathbb{E}(y|\mathbf{x}) = \mathbf{x}'\beta^*$ , 其中,  $\beta^*$  是  $K \times 1$  的系数向量。由 CEF 分解定理,

$$\mathbb{E}[\mathbf{x}(y - \mathbb{E}(y|\mathbf{x}))] = 0.$$

容易证明 :  $\beta^* = \beta$ . □

## The Best Linear Predictor Theorem (Regression Justification II)

The function  $\mathbf{x}'\boldsymbol{\beta}$  is the best linear predictor of  $y$  given  $\mathbf{x}$  in an MSE sense.

## The Regression CEF Theorem (Regression Justification III)

The function  $\mathbf{x}'\boldsymbol{\beta}$  provides the minimum MSE linear approximation to  $\mathbb{E}(y|\mathbf{x})$ , that is,

$$\boldsymbol{\beta} = \arg \min_{\mathbf{b}} \mathbb{E} \{ (\mathbb{E}(y|\mathbf{x}) - \mathbf{x}'\mathbf{b})^2 \}.$$

### Proof.

Write

$$\begin{aligned} (y - \mathbf{x}'\mathbf{b})^2 &= \{ (y - \mathbb{E}(y|\mathbf{x})) + (\mathbb{E}(y|\mathbf{x}) - \mathbf{x}'\mathbf{b}) \}^2 \\ &= [y - \mathbb{E}(y|\mathbf{x})]^2 + [\mathbb{E}(y|\mathbf{x}) - \mathbf{x}'\mathbf{b}]^2 + 2[y - \mathbb{E}(y|\mathbf{x})] \cdot [\mathbb{E}(y|\mathbf{x}) - \mathbf{x}'\mathbf{b}]. \end{aligned}$$

Therefore, minimize  $\mathbb{E} \{ (\mathbb{E}(y|\mathbf{x}) - \mathbf{x}'\mathbf{b})^2 \}$  is equivalent to minimize  $\mathbb{E} \{ (y - \mathbf{x}'\mathbf{b})^2 \}$ . □

# References



Joshua D. Angrist & Jörn-Steffen Pischke (2009)

*Mostly Harmless Econometrics: An Empiricist's Companion*, Chapter 3.

Princeton University Press, 2009.



Bruce E. Hansen. (2020)

*Econometrics*, Section 2.18.

<https://www.ssc.wisc.edu/bhansen/econometrics/> [▶ Link](#)